



F U N D A Ç ã O  
GETULIO VARGAS

**EPGE**

Escola de Pós-Graduação  
em Economia

Ensaio Econômico

Escola de

Pós-Graduação

em Economia

da Fundação

Getúlio Vargas

Nº 487

ISSN 0104-8910

## Using Irregularly Spaced Returns to Estimate Multi-Factor Models: Application to Brazilian Equity Data

Alvaro Veiga, Leonardo Rocha Souza

Junho de 2003

URL: <http://hdl.handle.net/10438/676>

Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getulio Vargas.

#### ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA

Diretor Geral: Renato Fragelli Cardoso

Diretor de Ensino: Luis Henrique Bertolino Braidó

Diretor de Pesquisa: João Victor Issler

Diretor de Publicações Científicas: Ricardo de Oliveira Cavalcanti

Veiga, Alvaro

Using Irregularly Spaced Returns to Estimate  
Multi-Factor Models: Application to Brazilian Equity Data/  
Alvaro Veiga, Leonardo Rocha Souza - Rio de Janeiro : FGV,EPGE,  
2010

(Ensaaios Econômicos; 487)

Inclui bibliografia.

CDD-330

**Nº 487**

ISSN 0104-8910

***Using irregularly spaced returns to estimate multi-factor models:  
Application to brazilian equity data***

***Álvaro Veiga  
Leonardo Rocha Souza***

*Junho de 2003*

Using Irregularly Spaced Returns to Estimate Multi-Factor Models:  
Application to Brazilian Equity Data

**Álvaro Veiga**

Departamento de Engenharia Elétrica – PUC Rio

**Leonardo R. Souza**

Graduate School of Economics (EPGE) – Fundação Getúlio Vargas

**Abstract** – Multi-factor models constitute a useful tool to explain cross-sectional covariance in equities returns. We propose in this paper the use of irregularly spaced returns in the multi-factor model estimation and provide an empirical example with the 389 most liquid equities in the Brazilian Market. The market index shows itself significant to explain equity returns while the US\$/Brazilian Real exchange rate and the Brazilian standard interest rate does not. This example shows the usefulness of the estimation method in further using the model to fill in missing values and to provide interval forecasts.

**Keywords:** Multi-Factor Model, Missing Data, Irregularly Spaced Returns

**JEL Classification:** C32, C53 e G13

## 1 - Introduction

Emergent markets frequently suffer from low liquidity and tend to concentrate most of transactions on a few liquid assets<sup>1</sup>. For instance, the Brazilian Equity Market comprises about 1190 different stocks but almost 40% have not been negotiated in the last year and 32% of the remaining has been negotiated less than once a month. As a consequence, for many stocks there will be no prices for a large proportion of days. Still worse, traditional missing value imputation methodologies do not take into account prices of transactions occurring in a given day but not in the previous and in the subsequent days. As they work with daily returns, they discard the information on prices of these “isolated” days, since the computation of daily returns require the existence of transactions in two subsequent days.

Financial institutions, however, do need those prices and returns everyday in order to fulfil regulatory requirements and implement their methodologies of quantitative analysis of risk and return<sup>2</sup>. Moreover, given some statistical properties of the data, widespread accurate pricing of non-traded equities diminishes arbitrage opportunities. A market model can provide the expected values of missing prices and returns. Sharpe (1964) proposed the Capital Asset Pricing Model (CAPM) to explain asset returns. However, a number of papers provided empirical evidence against the CAPM. For example, Bhandari (1988) and Chan, Hamao and Lakonishok (1991). Furthermore Fama and French (1992) and Jegadeesh (1992) show that the market beta has little power in explaining cross-sectional asset returns, meaning that some additional common “factor” could further explain the returns. Ross (1976) proposed the Arbitrage Pricing Theory (APT), which allowed more than one factor to explain the assets returns and consequently diversify risk premia. However, in all these works the factors are non-observable. Reinganum (1981) and Mei (1993) use autoregressive approach to explain the hidden factors. Chen, Roll and Ross (1986) introduced macroeconomic variables to explain monthly stock returns in a multi-factor linear regression. However, this approach is limited by the availability of macroeconomic data, in which many of the variables have frequency of observation not higher than once a month, while financial institutions need daily estimates. Even so, it is common practice in these institutions to use a similar model to explain assets returns, where the most used risk factors are

---

<sup>1</sup> Subramanian (2001, p.77), for example, observes this effect in bond markets.

<sup>2</sup> For an example in the Brazilian Market one can cite the resolution number 2804, which regulates liquidity risk, decreed by the Brazilian Central Bank in December 21, 2000.

Market indexes, foreign exchange rates and those related to interest rates. Then a circular problem may arise: the Market model must be estimated before computing the expected value of missing prices and returns, but many numerical problems may surge and bias the estimation if the proportion of missing data is high. A biased estimation will lead to poor price filling. Also, since the Market model uses returns as the dependent variable, instead of prices, the proportion of missing values will be greater than the proportion of missing prices. For instance, note that a stock that is negotiated every second day will generate no daily return at all.

In this paper, to circumvent the problem mentioned above, we propose an estimation method to the multi-factor model which makes use of irregularly spaced returns, enabling the use of every historical price available and thus increasing efficiency. Furthermore, we allow for the weights assigned to the data to decrease exponentially with age, so that the distant past does not influence much the estimates on the ever-changing market. Other authors have tried different approaches to estimate “betas” for infrequently traded assets. Scholes and Williams (1977), Dimson (1979), and Fowler and Rorke (1983) propose lagged regressions to have their OLS estimates combined in some way. Brooks et al. (2001, 2002) correct the beta OLS estimates for bias and inconsistency using a sample selectivity model. None of them, however, seem to have used irregular returns. Marsh (1979) proposes an estimation method using irregular returns to help testing some market hypothesis. His approach is similar to ours, but he does not take into account the age of the observations. Dimson and Marsh (1983) use the same approach as Marsh (1979), but they study the stability of the estimates across non-overlapping periods.

In a practical illustration, there are equities with as low as 5% of transaction days per year, for which the multi-factor model can be used to estimate missing prices and also to estimate the  $h$ -step ahead interval forecast for returns, conditioned on the risk factors (that must be taken into account exogenously). We work with  $h = 1, 2, 3$ , having as motivation to use up to three steps ahead the fact that the Sao Paulo Stock Exchange Market (Bovespa) liquidity system (represented by its clearinghouse, CBLC) determines that parties involved in a transaction have three days to liquidate it. To evaluate the estimation method, we ran a back test with the daily closing prices of 389 stock over the period ranging from August 25, 1998 to February 28, 2001. Two main important aspects of the estimated model were examined, both having practical implications to financial institutions. First, we examine its ability to produce good

estimation of missing prices. This is done comparing each observed price to the estimated value that would be produced by the model if this particular price was missing. Second, the  $h$  steps ahead interval forecasts are evaluated by comparing the nominal and observed frequencies of values in the 5% tails, which is an important measure of risk (Jorion, 1997). For comparative purposes, the same experiment is run to other commonly used imputation methodologies, such as repeating the last price (naïf method) or mimicking the market index return, as well as using the multi-factor model estimated only with the available daily returns. All these are nested under the same multi-factor class of models with possibly time-varying coefficients. We conclude that the multi-factor model estimated using irregular returns globally outperforms all the others, including the traditional regular returns estimation. Furthermore, the Brazilian Equity Market Index (IBV) shows to significantly explain equity returns but the US Dollar/ Brazilian Real exchange rate and the Brazilian inter-bank overnight interest rate do not.

The next Section presents the notation used further in this paper and the methodology proposed here, while Section 3 shows an empirical example with Brazilian equity data. Section 4 offers some concluding remarks.

## 2 – Notation and Methodology

In this Section we describe the notation and the methodology used to estimate the multi-factor model making use of irregularly spaced returns. The price of equity  $j$  at time  $t$  is denoted by  $P_t^j$  and the (log-)return of the same equity at time  $t$  by:

$$R_t^j = \ln\left(\frac{P_t^j}{P_{t-1}^j}\right) = \ln(P_t^j) - \ln(P_{t-1}^j), \quad (1)$$

where  $j = 1, \dots, J$  ( $J = 389$  in the experiment) and  $t = 1, \dots, T$ . In our formulation, using log-returns is necessary to preserve additivity of daily returns over  $\tau$ -days periods ( $\tau$  integer and greater than one).

### 2.1 - Multi-Factor Model

The usual Multi-Factor model with  $p$  factors describes a linear relation between the return of equity  $j$  and some market indicators, taking the form

$$R_t^j = (\beta_t^j)^T X_t + \varepsilon_t^j, \quad (2)$$

where  $\beta_t^j \in B \subset \mathbb{R}^{p+1}$ ,  $\beta_t^j = [\beta_{0t}^j, \dots, \beta_{pt}^j]$ , is the vector of coefficients of equity  $j$  at time  $t$ ,  $X_t = [1, X_{1t}, \dots, X_{pt}]$  is the vector of market indicators (possibly log-returns) at time  $t$  and  $\varepsilon_t^j$  errors supposedly iid,  $\varepsilon_t^j \sim N(0, \sigma_j^2)$ . We allow the coefficients to vary slowly and smoothly in time to let the relationship between returns and indicators be dynamic as believed to occur in real markets. Note that this assumption may not hold since the  $\beta$ 's may be influenced by the factors magnitude, analogously to a “leverage effect”. However, the study of this variation beyond the scope of the paper. To deal with this variation in a simple fashion, we estimate all models within the exponentially weighted moving average (EWMA) framework (J.P. Morgan, 1995), which is briefly explained in Section 2.3.

Given there is no missing value in the estimation window, the estimation is obtained via weighted least squares, with the weights exponentially decreasing with the age of the data. The missing value of equity  $j$  occurs when there is no trade of the equity  $j$  at time (day)  $t$ . Its imputation at time  $t$  based on this model uses the returns expected values given the current (daily) market indicators at time  $t$ . The VaR (Value at Risk, which “measures the worst expected loss over a given time interval under normal market conditions at a given confidence level”, Jorion, 1997, p. xiii) estimation  $h$  days ahead, however, requires forecasts (possibly density forecasts) of the market indexes.

## 2.2 – Irregularly Spaced Returns

Emergent markets contain a number of illiquid equities which are not negotiated every day, unlike the market indicators (the risk factors), which are available every business day. For these equities, the estimation window is scattered with missing values. We assume in this paper that the data are missing at random (MAR), whether or not this randomness is endogenous or exogenously implied by information<sup>3</sup> arrivals that are randomly distributed across time. The latter would incorporate microstructure effects as in Easley et al. (1996), for example, in the error term. There would be, thus, some fraction of information that is common to all stocks, which might be explained by the risk factors, and other fraction of information concerning only a single or a small group of stocks, which would be modelled primarily in the error term. Other works (e.g. Engle and Russell, 1998) model the duration (time between subsequent trades) as an autoregressive process and find empirical evidence of transaction clustering. These

---

<sup>3</sup> Possibly private information.

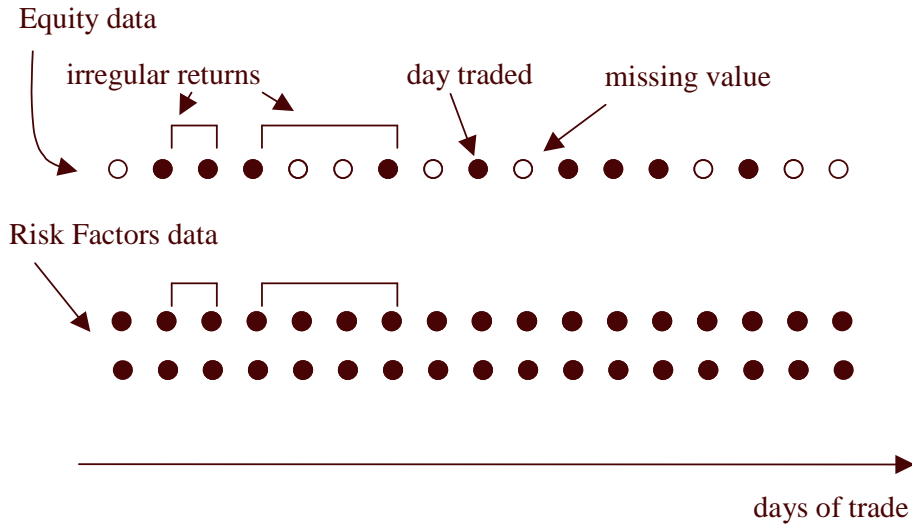


results, however, refer to high-frequency data (therefore liquid equities), whereas our paper refers to daily data.

Assume that equity  $j$  does not trade at time  $t$ . Lo and MacKinlay (1990) consider that the “observed return”  $R_t^j = 0$ , while there is a non-observed “virtual return”. They prove that, for an infrequently traded stock, spurious negative autocorrelations are present in the observed returns even if the price follows a random walk. In our framework there is only the “real” return, being it observed or not, coinciding with their “virtual return”. When the “real” return is not observed, it is just missing, so that one of our purposes is to make inferences on it. Using irregular returns to estimate the model and subsequently filling in missing prices avoids the effects of spurious autocorrelations, also found in empirical applications by Lo and MacKinlay (1988) and Atchison et al. (1987).

Let one consider the estimation window of a market model for a specific equity. Suppose the risk factors are observable every trading day, but there is a number of days in which the equity was not traded and thus no price was observed. Figure 1 illustrates how missing data determines the use of irregular returns.

**Figure 1:** Estimation window with missing data and irregular returns



Define the  $\tau$ -days return at time  $t$  for equity  $j$  as:

$$R_{t,\tau}^j \equiv \ln P_t^j - \ln P_{t-\tau}^j. \quad (3)$$

Moreover, equalling the right-hand sides of equations (1) and (2) and rearranging yields:

$$\ln P_t^j = \ln P_{t-1}^j + (\beta_t^j)^T X_t + \varepsilon_t^j. \quad (4)$$

Applying Equation (4) to  $t = t, t-1, t-2, \dots, t-\tau+1$  and summing up gives:

$$\ln P_t^j = \ln P_{t-\tau}^j + \sum_{i=0}^{\tau-1} (\beta_{t-i}^j)^T X_{t-i} + \sum_{i=0}^{\tau-1} \varepsilon_{t-i}^j. \quad (5)$$

If for practical purposes  $\beta_t^j$  and  $\sigma_{j,t}^2$  are considered (almost) constant on the interval  $[t-\tau, t]^4$ , and since  $\varepsilon_t^j$  is iid, then, using (3), (5) rewrites as:

$$R_{t,\tau}^j = (\beta_t^j)^T \sum_{i=0}^{\tau-1} X_{t-i} + w_{t,\tau}^j, \quad (6)$$

where

$$w_{t,\tau}^j = \sum_{i=0}^{\tau-1} \varepsilon_{t-i}^j; \quad w_{t,\tau}^j \sim N(0, \tau \sigma_j^2).$$

### 2.3 – Exponentially Weighted Moving Average (EWMA)

The EWMA framework has been widely used in practice since the RiskMetrics (J.P. Morgan, 1995) methodology was proposed. Its motivation is that the process generating the data may change smoothly through time. So, the older are the data the less weight it must have attached to. These weights decrease exponentially with time, according to the smoothing parameter  $\lambda$ ,  $0 < \lambda < 1$ . To day  $t^*$  is thus given weight  $\lambda$  times the weight given to the day  $t^*+1$ . The parameter  $K$ , in turn, determines the number of effective days to be used in the estimation window. The mean of equity  $j$  returns is then estimated at time  $t$  by:

$$\hat{\mu}_t^j = \sum_{i=0}^{K-1} \lambda^i (1-\lambda) R_{t-i}^j. \quad (7)$$

The covariances between returns of equities  $i$  and  $j$ , as well as the variance of equity  $i$  are estimated at time  $t$  by:

$$\begin{cases} \hat{\sigma}_{ij,t}^2 = C \sum_{m=0}^{K-1} \lambda^m (1-\lambda) (R_{t-m}^i R_{t-m}^j) - \hat{\mu}_t^i \hat{\mu}_t^j \\ \hat{\sigma}_{i,t}^2 = C \sum_{m=0}^{K-1} \lambda^m (1-\lambda) (R_{t-m}^i)^2 - (\hat{\mu}_t^i)^2 \end{cases}. \quad (8)$$

---

<sup>4</sup> There is no data in between  $t-\tau$  and  $t$  to estimate how  $\beta$  and  $\sigma^2$  change. Furthermore, if  $\tau$  is small the coefficients  $\beta$  and the variance  $\sigma^2$  are believed to change little.

where  $C$  is such that the sum of weights  $C\lambda^m(1-\lambda)$  is the unity ( $C \rightarrow 1$  as  $K \rightarrow \infty$ ). We do not take into account the loss of a degree of freedom in calculating the mean, but this has little effect since we use  $K = 252$  (corresponding to one year of data). It is usual to consider the mean of returns as being zero, simplifying more these equations. However, this assumption is not made here. An overview on EWMA is given in Alexander (1996).

## 2.4 – Methodology

### 2.4.1 – Estimation with Irregular Returns

Some equities have trading prices for all days in the estimation window. In these cases, Equation (2) is considered and a weighted least squares procedure, with EWMA weights, yields the estimates of  $\beta_t^j$  for each equity  $j$  at the current time  $t$ . This is done through the equation:

$$\hat{\beta}_t^j = (X^T W X)^{-1} X^T W R^j, \quad (9)$$

where

$R^j = R_{K \times 1}^j = [R_{t-K+1}^j \quad R_{t-K+2}^j \quad \dots \quad R_{t-1}^j \quad R_t^j]^T$  is the vector of returns of equity  $j$ ;

$$X = X_{K \times (p+1)} = \begin{bmatrix} (X_{t-K+1})^T \\ \vdots \\ (X_{t-1})^T \\ (X_t)^T \end{bmatrix} = \begin{bmatrix} 1 & X_{1,t-K+1} & \dots & X_{p,t-K+1} \\ 1 & X_{1,t-K+2} & \dots & X_{p,t-K+2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1,t} & \dots & X_{p,t} \end{bmatrix} \text{ is the design matrix (the same}$$

$$\text{for all equities); and } W = W_{K \times K} = \begin{bmatrix} \lambda^{K-1} & 0 & \dots & 0 \\ 0 & \lambda^{K-2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix} \text{ is the weights matrix}^5. \text{ In this}$$

case,  $p$  market indicators are considered.

Note that the lines of  $R^j$ ,  $X$  and  $W$  correspond to one day return. However, if there is a missing value in the estimation window, a line will correspond to a  $\tau$ -days return since (6), instead of (2), is considered. The redefined matrices  $R^j$ ,  $X$  and  $W$ , where all lines corresponding to a same  $\tau$ -days return collapse into one single line, are as follows:

---

<sup>5</sup> In fact, if all the weights in  $W$  are multiplied by the same constant, the estimates do not change, as the change in the inversion of  $X^T W X$  compensates the change in  $X^T W R^j$ . That is why the weights do not match with those in 2.3.

$$\begin{aligned}
R^j = R_{K_j \times 1}^j &= \begin{bmatrix} R_{t-i_1, \tau_1}^j \\ R_{t-i_2, \tau_2}^j \\ \vdots \\ R_{t-i_{K_j}, \tau_{K_j}}^j \end{bmatrix}; \quad X = X_{K_j \times (p+1)} = \begin{bmatrix} \tau_1 & \sum_{m=0}^{\tau_1-1} X_{1, t-i_1-m} & \cdots & \sum_{m=0}^{\tau_1-1} X_{p, t-i_1-m} \\ \tau_2 & \sum_{m=0}^{\tau_2-1} X_{1, t-i_2-m} & \cdots & \sum_{m=0}^{\tau_2-1} X_{p, t-i_2-m} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{K_j} & \sum_{m=0}^{\tau_{K_j}-1} X_{1, t-i_{K_j}-m} & \cdots & \sum_{m=0}^{\tau_{K_j}-1} X_{p, t-i_{K_j}-m} \end{bmatrix} \\
W = W_{K_j \times K_j} &= \begin{bmatrix} \frac{\lambda^{i_1}}{\tau_1} & 0 & \cdots & 0 \\ 0 & \frac{\lambda^{i_2}}{\tau_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\lambda^{i_{K_j}}}{\tau_{K_j}} \end{bmatrix}, \tag{10}
\end{aligned}$$

where there are  $K_j+1$  prices ( $K_j$  possibly irregularly spaced returns) available in the estimation window for equity  $j$ ;  $\tau_k$ ,  $k = 1, 2, \dots, K_j$ , is the interval of the  $k$ -th return in the estimation window; and  $i_k$  is the time past between the  $k$ -th return of equity  $j$  and the current time  $t$  (in other words, the age of the return, measured in days). So, in addition to account for exponentially decaying weights,  $W$  assigns weights inversely proportional to the variances of the errors  $w_{t, \tau}^j$ <sup>6</sup>.

#### 2.4.2 – Missing Values Imputation

In the more general case, the imputation of missing values can occur at any point of the estimation window, although the results with real data shown here are only from imputation at the end of it (current day). For the sake of simplicity, consider  $p_t = \ln P_t^j$ . Following (5), the expected value of  $p_t$ , conditional on  $\beta_t^j$ , the matrix of market indicators  $\tilde{X}_{t+\tau_2} = [X_1, X_2, \dots, X_t, \dots, X_{t+\tau_2}]$  and the next existing price (forward or backward respectively for equations 11.A and 11.B) is given by:

$$E(p_t / p_{t-\tau_1}, \beta_t^j, \tilde{X}_{t+\tau_2}) = p_{t-\tau_1} + (\beta_t^j)^T \sum_{i=0}^{\tau_1-1} X_{t-i} \quad \text{and} \tag{11.A}$$

$$E(p_t / p_{t+\tau_2}, \beta_t^j, \tilde{X}_{t+\tau_2}) = p_{t+\tau_2} - (\beta_t^j)^T \sum_{i=0}^{\tau_2-1} X_{t+\tau_2-i}, \tag{11.B}$$

with respective variances:

---

<sup>6</sup> In order to minimize the estimator variance (see, for example, Neter et al, 1996, p.400).

$$\text{Var}(p_t / p_{t-\tau_1}, \beta_t^j, \tilde{X}_{t+\tau_2}) = E\left(\left[\sum_{i=0}^{\tau_1-1} \varepsilon_{t-i}\right]^2\right) = \tau_1 \sigma_j^2 \text{ and} \quad (12.A)$$

$$\text{Var}(p_t / p_{t+\tau_2}, \beta_t^j, \tilde{X}_{t+\tau_2}) = \tau_2 \sigma_j^2 \quad . \quad (12.B)$$

Using (11.A) and (11.B), and replacing  $\beta_t^j$  by its estimate  $\hat{\beta}_t^j$ , we obtain the final expression for the estimates of a missing price.

$$\hat{p}_{t/\tau_1} = p_{t-\tau_1} + (\hat{\beta}_t^j)^T \sum_{i=0}^{\tau_1-1} X_{t-i}, \quad (13.A)$$

$$\hat{p}_{t/\tau_2} = p_{t+\tau_2} - (\hat{\beta}_t^j)^T \sum_{i=0}^{\tau_2-1} X_{t+\tau_2-i}. \quad (13.B)$$

If the missing price is before the first observed price, then (13.B) is used. On the other hand, if the missing price is after the last observed price, (13.A) is used. Otherwise, when the missing price has observed prices both before and after it, the minimum variance combination of (11.A) and (11.B) is estimated by:

$$\hat{p}_t = \gamma \hat{p}_{t/\tau_2} + (1-\gamma) \hat{p}_{t/\tau_1} = \frac{\tau_1}{\tau_1 + \tau_2} \hat{p}_{t/\tau_2} + \frac{\tau_2}{\tau_1 + \tau_2} \hat{p}_{t/\tau_1}. \quad (14)$$

Note that we do not take into account the uncertainty on the estimated coefficients in (13.A) and (13.B) to minimize the variance in (14). On the contrary, the minimization is performed in (11.A) and (11.B), simplifying considerably the calculation. For details on the minimum variance estimator see Neter et al (1996, p.400).

### 2.4.3 – Estimation of Interval Forecasts

Factor models are not tailored to produce forecasts since they will depend on forecasts of the risk factors themselves. In this paper, we consider the interval forecast (IF) estimation as an *ad hoc* procedure, ignoring uncertainties of many sorts and making some further simplifying assumptions. For example, the exact parametric predictive density given the specified model would require the account of the variance of estimated coefficients, even if the assumptions hold, and this is not considered.

If one is to consider the problem through the view of the clearinghouse, which guarantees the transactions, two-sided IF's must be supplied. This is because any of the two parties may default, which means that the clearinghouse may have to buy or sell the equity in the market three days after the transaction is agreed. The IF is computed as the

following. Let  $E(R_{t+\tau,\tau})$  be the expected value of equity  $j$   $\tau$ -days return at time  $t+\tau$  given  $\beta_t^j$  and  $\tilde{X}_{t+\tau}$ :

$$E(R_{t+\tau,\tau}^j | \beta_t^j, \tilde{X}_{t+\tau}) = E\left(\sum_{i=1}^{\tau} (\beta_{t+i}^j)^T X_{t+i} + w_{t,\tau}\right) \quad (15)$$

It is straightforward to estimate the  $\tau$ -days ahead return at time  $t$  by:

$$\hat{R}_{t+\tau,\tau}^j = (\hat{\beta}_t^j)^T \sum_{i=1}^{\tau} \hat{X}_{t+i} \quad (16)$$

Note that (16) depends on forecasts of the risk factors themselves. However, the method used to forecast the risk factors is beyond the scope of the paper. Now we estimate the variance  $\tau$ -days ahead. We make a simplifying assumption that the risk factors are homoskedastic, or at least that their variances vary slowly in the  $\tau$ -days ahead period. Alternatively to making this assumption, this variance may be viewed as conditional on the current circumstances concerning volatility and risk. Since it was assumed before that the coefficients  $\beta$  are (almost) constant and the errors are serially independent, the variance of equity  $j$   $\tau$ -days return at time  $t+\tau$  may be approximated by:

$$\sigma_{R_{t+\tau,\tau}^j}^2 \cong \tau \left[ \sigma_{j,t}^2 + \beta_t^{jT} \Sigma_{X_t} \beta_t^j \right], \quad (17)$$

where  $\Sigma_{X_t} = E[(X_t - E[X_t])(X_t - E[X_t])^T]$ . Ignoring the uncertainty caused by the estimation of the coefficients  $\beta$ , this variance can be estimated at time  $t$  by:

$$\hat{\sigma}_{R_{t+\tau,\tau}^j}^2 = \tau \left[ \hat{\sigma}_{j,t}^2 + \hat{\beta}_t^{jT} \hat{\Sigma}_{X_t} \hat{\beta}_t^j \right]. \quad (18)$$

The risk factors variance-covariance matrix (VCV) must be estimated and there are a number of methods to do so. Alexander and Leigh (1997) study the accuracy of some of these methods considering the proportion of returns that fall below the estimated VaR. The IF must then be based on a predictive distribution, which exact form is non-trivial to obtain. If the betas were known, as well as the future values of  $X_t$ , the conditional distribution of the forecast errors  $R_{t+\tau,\tau}^j - \hat{R}_{t+\tau,\tau}^j$  would be a Student's  $t$  with  $K_j - p - 1$  degrees of freedom multiplied by  $\hat{\sigma}_{j,t}$ , but the betas are estimated and the risk factors forecast. Nonetheless, this parametric form will be used in place of the exact one, just incorporating the risk factors uncertainty in the variance as in (18). The lower and upper bounds for a  $(1-\alpha)100\%$  IF are given then by:

$$L = \exp\left[\ln(P_t^j) + \hat{R}_{t+\tau,\tau}^j + t^{-1}(\alpha/2, K_j - p - 1)\hat{\sigma}_{R_{t+\tau,\tau}^j}\right] \text{ and}$$

$$U = \exp\left[\ln(P_t^j) + \hat{R}_{t+\tau,\tau}^j + t^{-1}(1-\alpha/2, K_j - p - 1)\hat{\sigma}_{R_j,t+\tau,\tau}\right], \quad (19)$$

where  $t^{-1}(\cdot, v)$  is the inverse cumulative distribution function of a Student's  $t$  random variable with  $v$  degrees of freedom, and  $K_j + 1$  is the number of observed prices of equity  $j$  in the estimation window. The lower bound, if desired, can be calibrated using the semi-variance (see Gastineau e Kristzman, 1996, p. 250, for example) to estimate  $\hat{\sigma}_{j,t}^2$ . The semi-variance is the mean of the squared **negative** deviations from the mean, and its inspiration comes from the asymmetry observed in financial returns. The VaR at  $1-\alpha/2$  confidence level is given by the lower bound of the  $1-\alpha$  IF.

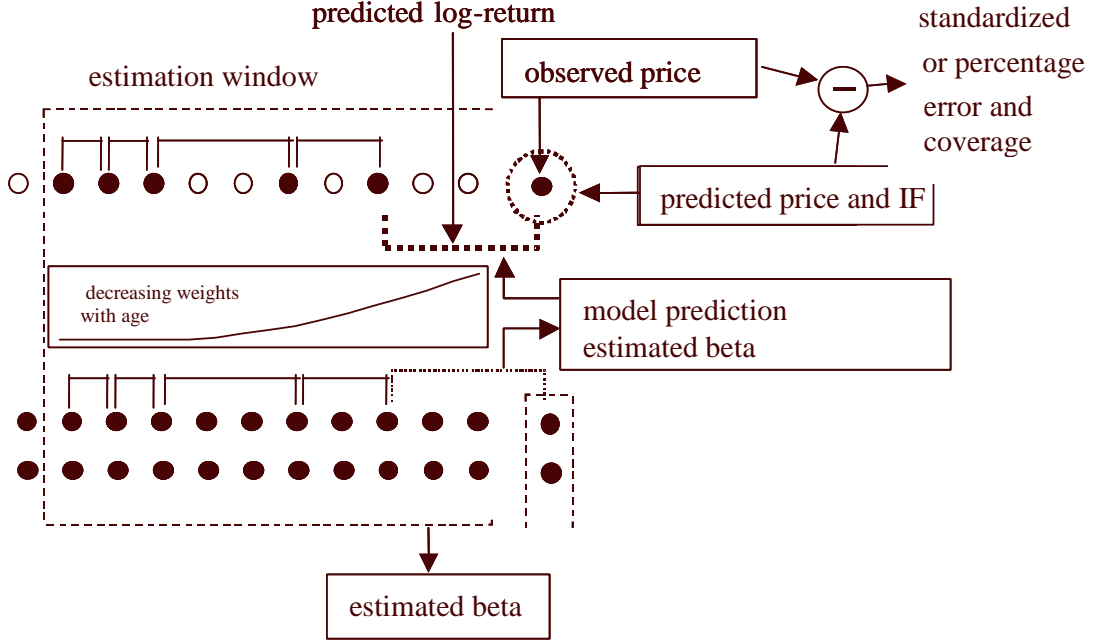
### 3 – Empirical Example with Brazilian Equity Data

The database consists of 389 stock closing prices (the most liquid in the Brazilian Market) over the period ranging from August 25, 1998 to February 28, 2001. Considering closing prices as daily prices (i.e., regularly spaced) is an approximation, since the last trade needs not to be near the end of the trading period. However, the time in the day at which each equity's last transaction occurs is unavailable to us. The results are divided in two sections, the first dedicated to the imputation of missing values and the second to the IF estimation. To access the accuracy of the missing values imputation, each single observation in the database within specified patterns is deleted at a time and filled in as is schematised in Figure 2. This enables the computation of error statistics on the imputation. One might argue that a sample selection bias may permeate this kind of comparison<sup>7</sup> in the case days with missing data are structurally different from days with trade, using microstructures argumentation. This would make it impossible to disentangle from the sample selection bias when evaluating the performance of missing values imputation methodologies for equity prices, since one would be left nothing to compare the price estimates with. On the other hand, any significant correlation found between the risk factors and the equity returns would allow one to anticipate future returns of this equity, in such a way that by a no arbitrage point-of-view our methodology is correct.

---

<sup>7</sup> The authors thank Daniel Ferreira for raising this concern in a seminar presentation at the Getulio Vargas Foundation.

Figure 2: Schematisation of the experiment.



In turn, the IF is estimated for every observation in the database within the same specified patterns, one, two and three steps ahead, and its coverage verified. The specified patterns are prices available at  $(t, t+1)$ ,  $(t, t+2)$  and  $(t, t+3)$ . These patterns are motivated by the fact that the parties involved have three days to liquidate the transaction, and thus a clearinghouse must estimate a IF for the equity price up to three days ahead in order to access its risk in the case of default.

### 3.1 – Missing Values Imputation

The imputation of missing values must be performed in the end of each day, so that the values of market indexes in the irregular returns factor model are known up to  $t$ . The accuracy of the imputation given by the proposed method is compared with the following methods: the *naïf*, where the last price available is repeated to fill in a missing value (predicted returns are zero); and the market *proxy* (the São Paulo Stock Exchange Market Index), where the equity is supposed to follow the market return in the absence of the real price. Note that both methods are nested in the multi-factor (*naïf*:  $\beta_{i,t}^j = 0$ ,  $i = 0, 1, \dots, p$ ; *proxy*:  $\beta_{k,t}^j = 1$ , where  $\beta_{k,t}^j$  is the coefficient referring to the market index return, and  $\beta_{i,t}^j = 0$ ,  $i \neq k$ ). A further comparison includes the multi-factor estimated conventionally, that is, only with regular (daily) returns. In a companion paper, Souza



and Veiga (2001) implemented an E-M algorithm (Dempster, Laird and Rubin, 1977) with principal component analysis to compete against the multi-factor with irregular returns on the same database. Their conclusion is that the multi-factor outperforms the E-M for the equities with less than 95% of data available.

To find the best multi-factor configuration, we test three values of the EWMA smoothing constant,  $\lambda = 0.98, 0.99$  and  $1$ . These high values of  $\lambda$  are justified by the size of the estimation window (252 business days, approximately one year of data), which in turn is justified by the inclusion of many low liquidity equities in the comparison<sup>8</sup>.

The factors used to explain the equities returns were the Brazilian Market Index (IBV), based on the most liquid equities negotiated in the São Paulo Stock Market; the Brazilian inter-bank overnight interest rate (CDI); and the US Dollar/Brazilian Real exchange rate (US\$). Different combinations of these three factors were compared against each other. The comparison is done by means of the following statistics:

**RMSSE:** root mean square of standardized errors

$$RMSSE(\tau) = \sqrt{\frac{1}{N} \sum_{t,j} \left[ \frac{(\hat{R}_{t,\tau}^j - R_{t,\tau}^j)}{\hat{\sigma}_{R^j,t,\tau}} \right]^2}, \quad (20)$$

where  $N$  is the total number of cases of each pattern specified above across time and equities and  $\tau$  is the time horizon,  $\tau = 1, 2, 3..$  In order to keep the comparison fair,  $\hat{\sigma}_{R^j,t,\tau}$  is estimated by EWMA and is the same for all methods. Errors are standardized by the volatility estimate because some equities are more volatile than others, and so their prediction errors can be compared without some equities dominating the statistic in spite of others.

**DC:** direction of change statistic

$$DC(\tau) = \frac{1}{N} \sum_{t,j} \text{sgn}(\hat{R}_{t,\tau}^j) * \text{sgn}(R_{t,\tau}^j), \quad \text{where } \text{sgn}(a) = \begin{cases} 1, & \text{if } a > 0 \\ 0, & \text{if } a = 0 \\ -1, & \text{if } a < 0 \end{cases} \quad (21)$$

This statistic is related to the proportion of times the predicted return has the same sign than the actual return (predict the equity price will rise and it indeed rises or

---

<sup>8</sup> Some require one year to have, say, 6 or 10 days in which they are negotiated.

predict it will fall and it falls). It measures the difference between the number of times the method predicts the direction of change correctly and the number of times the direction of change is predicted incorrectly, relatively to the total number of cases.

**MAPE**: mean absolute percentage error

$$MAPE(\tau) = \frac{1}{N} \sum_{t,j} \left| \frac{\hat{P}_{t,\tau}^j - P_t^j}{P_t^j} \right|. \quad (22)$$

While the RMSSE measures the error in the return prediction, the MAPE measures the percentage error in imputed prices.

Tables 1 and 2 (one day ahead) and Figures 3-5 (three days ahead) show the results for the *naïf*, the proxy and the multi-factor with irregular returns using the following factors: no factor – only the constant (Of), the market index return (IBV), IBV and the one day standard interest rate (IBV & CDI), and the IBV and the return of the US\$/Brazilian Real exchange rate (IBV & US\$). All the multi-factor results refer to  $\lambda = 0.99$ . The results referring to the remaining values of  $\lambda$ , 0.98 and 1, are not shown as they are in general worse than those with  $\lambda = 0.99$ . The results are grouped by the adjusted  $R^2$  (of IBV) as it was the feature that most explained the difference between methods.

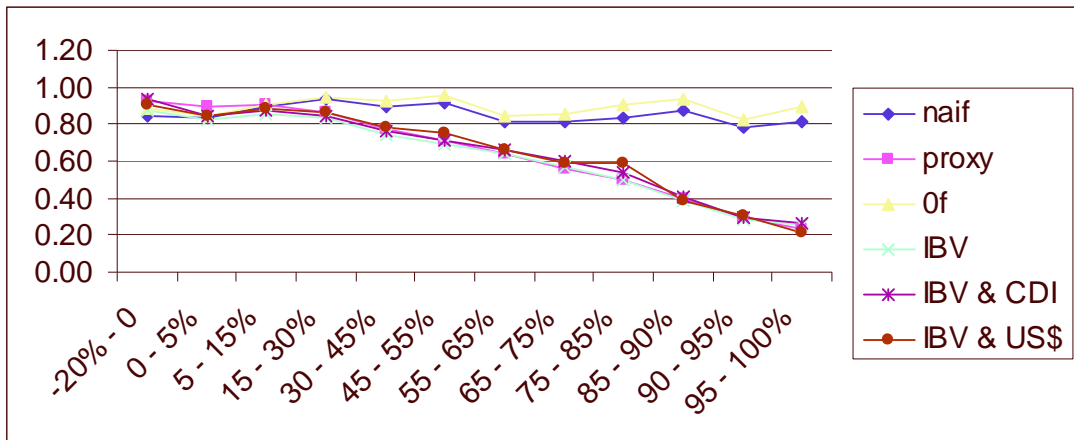
**Table 1:** RMSSE one day ahead by adjusted  $R^2$ .

adj. R2	naïf	proxy	Of	IBV	IBV & CDI	IBV & US\$	# cases
-20% - 0	<b>1.03</b>	1.12	1.04	<b>1.03</b>	1.05	1.05	6339
0 - 5%	0.96	1.01	0.97	<b>0.95</b>	0.96	0.96	23839
5 - 15%	0.94	0.96	0.94	<b>0.90</b>	0.91	0.91	31151
15 - 30%	0.93	0.88	0.93	<b>0.84</b>	<b>0.84</b>	0.85	24975
30 - 45%	0.87	0.77	0.88	<b>0.73</b>	0.74	0.74	13053
45 - 55%	0.89	0.71	0.90	<b>0.67</b>	0.68	0.71	5339
55 - 65%	0.81	0.63	0.82	<b>0.61</b>	0.62	0.63	3332
65 - 75%	0.79	0.55	0.80	<b>0.54</b>	0.55	0.55	1679
75 - 85%	0.77	0.47	0.79	<b>0.46</b>	0.47	0.48	1070
85 - 90%	0.83	0.42	0.86	<b>0.39</b>	<b>0.39</b>	<b>0.39</b>	440
90 - 95%	0.77	0.30	0.77	<b>0.29</b>	<b>0.29</b>	<b>0.29</b>	343
95 - 100%	0.73	<b>0.19</b>	0.77	0.20	0.21	0.20	141
all	0.88	0.85	0.90	<b>0.81</b>	0.83	0.84	111701

**Table 2:** MAPE one day ahead by adjusted  $R^2$ .

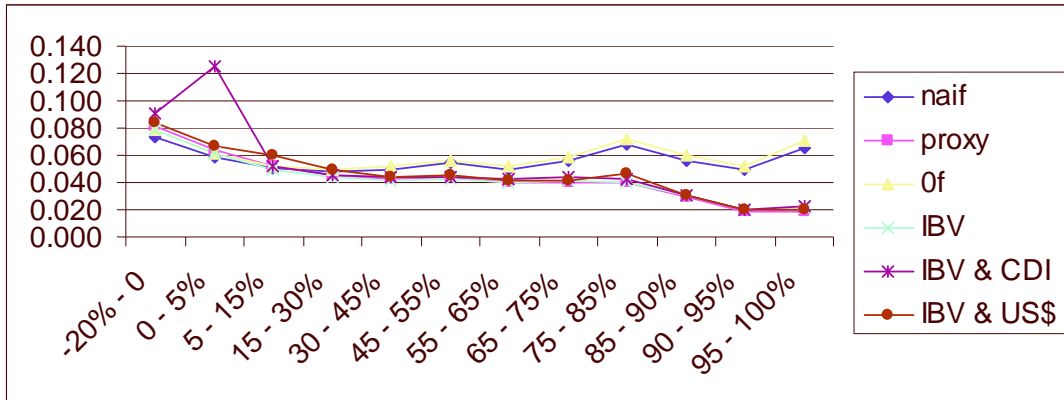
adj. $R^2$	naïf	proxy	Of	IBV	IBV & CDI	IBV & US\$
-20% - 0	<b>0.048</b>	0.054	0.050	0.051	0.054	0.052
0 - 5%	<b>0.038</b>	0.042	0.039	0.040	0.041	0.041
5 - 15%	<b>0.030</b>	0.032	0.031	<b>0.030</b>	0.031	0.031
15 - 30%	0.028	0.027	0.028	<b>0.026</b>	<b>0.026</b>	0.027
30 - 45%	0.028	0.025	0.028	<b>0.024</b>	<b>0.024</b>	0.025
45 - 55%	0.029	0.024	0.030	<b>0.023</b>	0.024	0.024
55 - 65%	0.029	<b>0.023</b>	0.030	<b>0.023</b>	<b>0.023</b>	0.024
65 - 75%	0.031	<b>0.022</b>	0.032	0.023	0.023	0.023
75 - 85%	0.037	<b>0.021</b>	0.038	0.022	0.022	0.023
85 - 90%	0.033	0.018	0.034	<b>0.017</b>	<b>0.017</b>	<b>0.017</b>
90 - 95%	0.028	<b>0.011</b>	0.029	<b>0.011</b>	0.012	0.012
95 - 100%	0.040	<b>0.010</b>	0.041	0.011	0.011	0.011
all	0.0320	0.0326	0.0328	<b>0.0308</b>	0.0315	0.0317

The RMSSE and the MAPE show that overall the multi-factor with irregular returns using only the market index returns performs best. This means that in general the market index is significant to explain equity returns, whereas the US\$/Brazilian Real exchange rate and the Brazilian inter-bank overnight interest rate are not so. However, the reader must keep in mind that 50 of these equities form the index, so that the experiment is biased in favour of this index (each of the two assets with most weight make up around 10% of the index, while the following five make up between 4% and 6% each<sup>9</sup>). On the other hand, the majority of these equities have a negligible weight in the index.

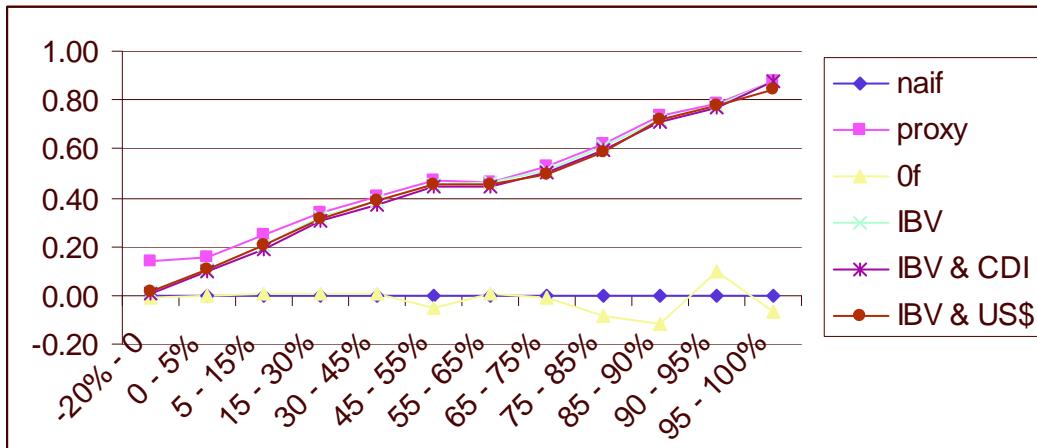
**Figure 3:** RMSSE three days ahead by adjusted  $R^2$ .

<sup>9</sup> Source: Algorithmics Brazil (<http://www.algorithmics.com.br/servicos/risco/AtivosIbovespa.asp>).

**Figure 4:** MAPE three days ahead by adjusted  $R^2$ .



**Figure 5:** DC three days ahead by adjusted  $R^2$ .



The naïf outperforms the multi-factor by slight margin where the adjusted  $R^2$  is low, and so does the proxy where the  $R^2$  is high. However, as the  $R^2$  begins to rise the naïf tends to be outperformed by all others, and furthermore the proxy is the worst method for low  $R^2$  (which is within the expected since the  $R^2$  is related to the explication coefficient for the IBV only, and the proxy uses the IBV with  $\beta = 1$ ). The DC statistic points to the proxy (the market index return) as the best indicator for the rise or fall of each equity price three days ahead. The results for one, two (not shown) and three days ahead are qualitatively similar.

### 3.1.1 – Irregular Versus Regular Returns

In this Section, the irregular returns multi-factor model is compared with the conventional multi-factor with daily returns. The simple fact that the conventional multi-factor needs two subsequent days of trade to provide a daily return compares

favourably to the multi-factor with irregular returns. The regular returns multi-factor discards the information of single prices (with no negotiation of the equity in the previous or in the next day) and part of the information brought by a price in the end of a block of prices, whilst the irregular returns version uses all prices. For this reason, there can be equities whose coefficients  $\beta$  can be estimated by the irregular returns version but not conventionally<sup>10</sup>. As long as there are degrees of freedom enough to reasonably estimate the coefficients  $\beta$ , the irregular returns multi-factor can be used, and in the present paper we considered 10 returns (less than 4% of data available, considering a 252 days estimation window) as a lower bound to estimate the regression. Table 3 compares the number of cases in the database for which there were enough data so that each version could estimate the coefficients  $\beta$ . Note that the regular returns enable the estimation only one third of the times the irregular returns do when there are at most 5% of data (prices) available, and approximately half of the times when there are between 5% and 15% of data available. From 65% of data available on both enable the same number of estimations in the database. The comparison reported below takes into account only the cases where both versions were able to estimate the coefficients. Since we stipulated 10 returns as a minimum to estimate the regression, the case where between 0% and 5% of the prices are available is restricted to 11 or 12 prices existing in the 252 days estimation window as can be seen in Table 3. In view of this, 33 cases where 10 daily returns could be computed out of 100 where 11 or 12 prices were available may seem too many, as one would expected existing prices scattered randomly over the estimation window. However, there are many cases in the database where a liquid equity started being negotiated during the period under study, having less than 13 prices in one year past because it was less than 13 days old then, justifying the high proportion.

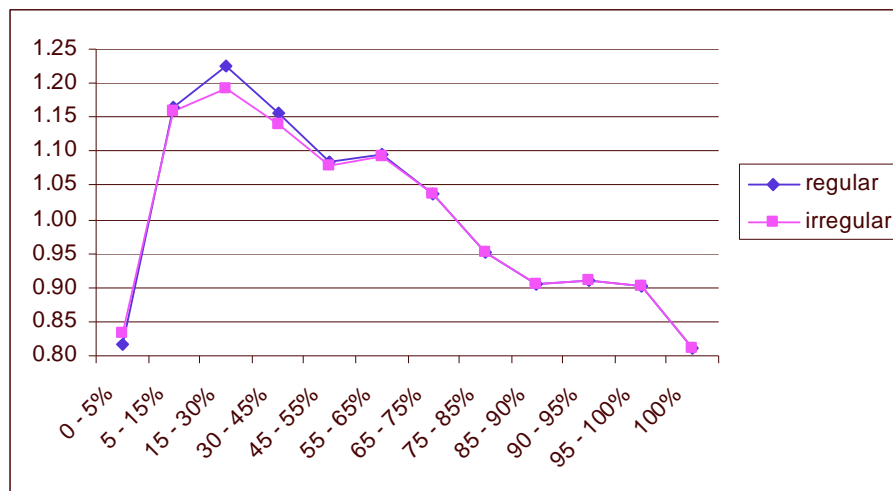
**Table 3:** frequency of estimation windows with more than 10 returns available (with an existing price in the end), for regular and irregular returns multi-factor. The cases are grouped by percentage of existing prices within the window.

	0 - 5%	5 - 15%	15 - 30%	30 - 45%	45 - 55%	55 - 65%	65 - 75%	75 - 85%	85 - 90%	90 - 95%	95 - 100%	100%
regular	33	824	3628	6073	5027	5976	7049	10260	7262	9131	24248	30822
irregular	100	1446	4006	6278	5097	6003	7049	10260	7262	9131	24248	30822

<sup>10</sup> In a degenerate case, an equity can be negotiated each second day, having thus availability of 50% of the daily prices but unavailability of daily returns. Even if one changes the frequency of observation, this would not be useful in the majority of cases, where the time between days in which the equity is traded is variable.

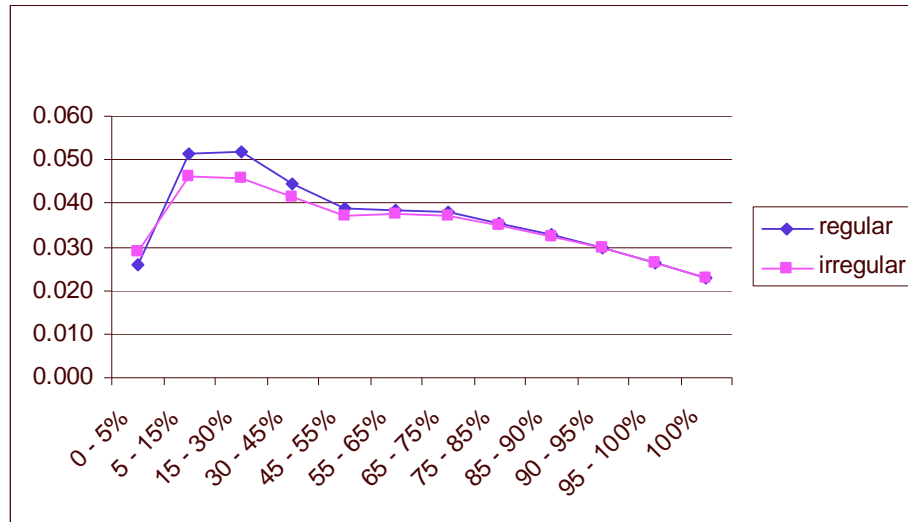
Figures 6 – 9 show the RMSSE and the MAPE for filled in data 1 and 3 days ahead. Note that the results are now grouped by percentage of existing prices in the estimation window, unlike the comparison with the naïf and the proxy. The irregular returns version yields almost always better results. The greatest difference appears when there are between 5% and 55% of prices available for imputation 1 day ahead<sup>11</sup> and between 0% and 75% of prices available for imputation 3 days ahead. When the horizon increases to 3 days ahead, the advantage of the irregular returns version over the conventional one is more apparent. When all the prices are available, both versions are equal by definition and so are their results. This also means that their results approximate more the more data exist in the estimation window.

**Figure 6:** RMSSE of filled in data from regular and irregular returns multi-factor, 1 day ahead, grouped by percentage of existing prices in the estimation window.

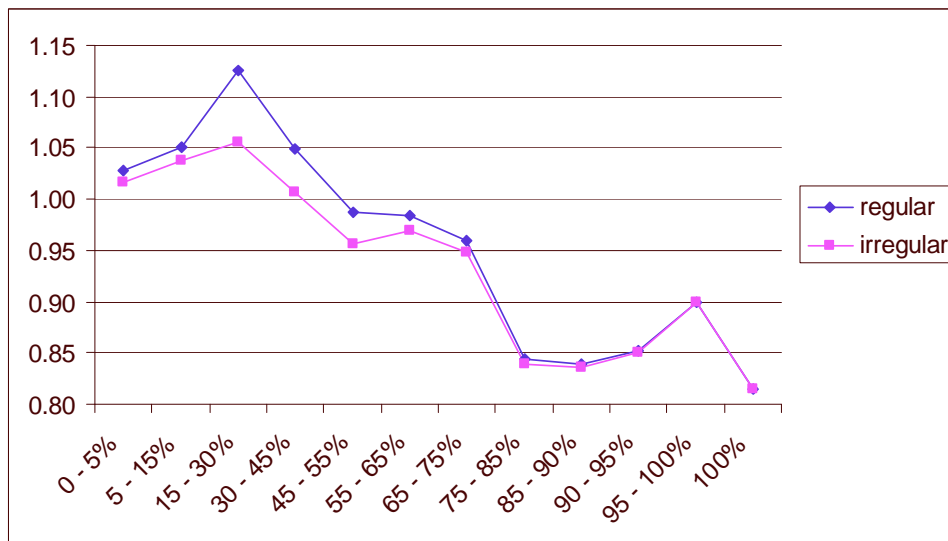


<sup>11</sup> Between 0% and 5% of prices available and 1 day ahead, the regular returns performed better, but note that this result is based on only 33 cases and that there were other 67 cases where only the irregular returns version could be used that were left out of the comparison.

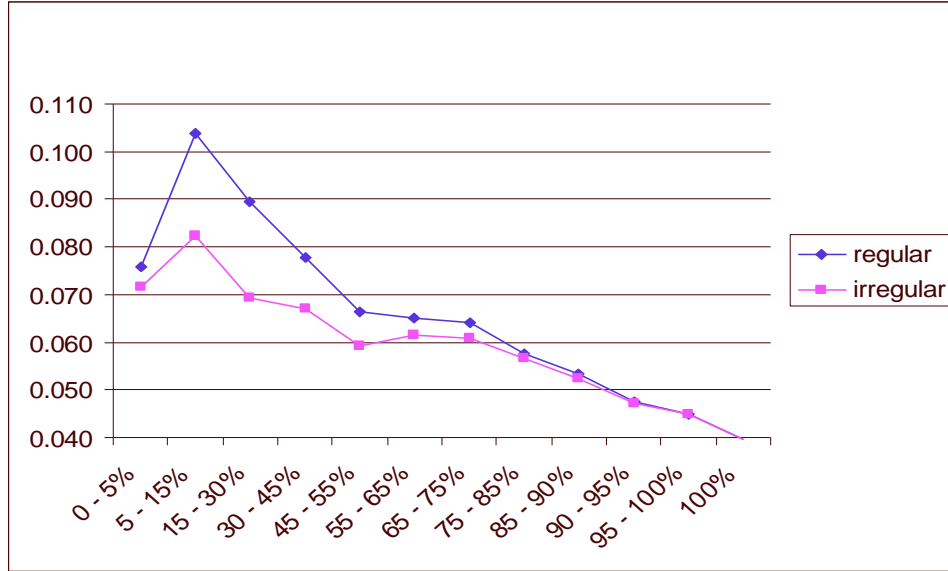
**Figure 7:** MAPE of filled in data from regular and irregular returns multi-factor, 1 day ahead, grouped by percentage of existing prices in the estimation window.



**Figure 8:** RMSSE of filled in data from regular and irregular returns multi-factor, 3 days ahead, grouped by percentage of existing prices in the estimation window.



**Figure 9:** MAPE of filled in data from regular and irregular returns multi-factor, 3 days ahead, grouped by percentage of existing prices in the estimation window.



It is clear by these results that the multi-factor with irregular returns outperforms the conventional multi-factor, in addition to enable the estimation in cases where the conventional cannot be applied. As pointed out before, as the proportion of existing data approaches the unity, both estimation methods (with only regular and with irregular returns) are more similar and hence have more similar results.

### 3.2 – IF estimation

In Section 3.1, it is clearly shown that the irregular multi-factor outperforms the naïf in filling in missing values, especially for higher values of the adjusted  $R^2$ . The proxy is slightly outperformed, especially for lower values of that statistic. From the previous results, we chose the best configuration to be the one that uses only the IBV as a factor. The 90% two-sided IF estimates (obtained via equation (19)) from this configuration are tested in this Section, together with a benchmark. The benchmark is the use of plain EWMA ( $\lambda = 0.99$ ) to estimate the IF of the equity price, under the normal assumption. The IF estimates from equation (19), however, need estimates for the risk factors volatilities. A volatility proxy using the highest and the lowest prices (assuming the log-price is a Brownian Motion) is tried (denoted by “irreg MF u-d” in the Tables 4 and 5), as well as the squared returns (denoted by “irreg MF”). Both were computed using EWMA with  $\lambda = 0.94$  (other values were tried and yielded no better results). The



volatility proxy using the highest and lowest prices is taken from Parkinson (1976) and is given by:

$$\hat{\sigma}^2 = \frac{(u - d)^2}{4 \ln 2}, \quad (23)$$

where  $u$  and  $d$  are the high and low prices normalized by the closing price of the previous day. A number of volatility proxies using the highest, the lowest, the opening and the closing prices are found in Garman and Klass (1980). All volatility proxies in the experiment use only data up to  $t-1$  to estimate the volatility at time  $t$ . We also tried the semi-variance but the results were no better and are not shown.

The percentage of cases the price fell below (above) the 90% central IF is shown in Tables 4 and 5, grouped by adjusted  $R^2$ . Starred and double starred numbers are significantly different from the nominal percentage of 5% at respective confidence levels of  $\alpha = 0.05$  and  $\alpha = 0.01$ . The significance was obtained using the likelihood ratio test of unconditional coverage proposed by Christoffersen (1998). However, it is important bear in mind that the higher number of observations, the higher is the test power to detect deviations from the nominal tail percentages, even if these deviations are so small that are irrelevant in practice. The great amount of data explains the high number of rejections, including those of coverages that are reasonable in practice (check, for instance, the first four lines of Tables 4 and 5). The main objective of this exercise is to get reasonable coverages. Indeed, the number of ignored sources of uncertainty does not let us postulate that the IF yields the exact coverage even if the model is correct.

**Table 4:** Percentage of cases below or above the 90% one day ahead central IF grouped by adjusted  $R^2$ .

Adj. R2	Below 90% IF			Above 90% IF			# cases
	irreg MF	irreg MF u-d	EWMA	irreg MF	irreg MF u-d	EWMA	
-20% - 0%	0.042**	0.042**	0.044	0.065**	0.065**	0.063**	5540
0 - 5%	0.037**	0.036**	0.039**	0.058**	0.057**	0.056**	24339
5 - 15%	0.033**	0.032**	0.037**	0.056**	0.056**	0.052	31820
15 - 30%	0.034**	0.036**	0.033**	0.055**	0.058**	0.053	26063
30 - 45%	0.033**	0.038**	0.030**	0.047	0.053	0.044**	12705
45 - 55%	0.037**	0.044	0.034**	0.051	0.059**	0.047	5058
55 - 65%	0.030**	0.041*	0.023**	0.044	0.053	0.041*	2922
65 - 75%	0.034**	0.044	0.026**	0.040	0.053	0.037*	1611
75 - 85%	0.038	0.056	0.025**	0.037	0.053	0.033*	885
85 - 90%	0.051	0.071	0.033	0.054	0.063	0.036	336
90 - 95%	0.027*	0.043	0.013**	0.040	0.064	0.030	299
95 - 100%	0.035	0.052	0.026	0.044	0.096*	0.043	115
all	0.035**	0.036**	0.035**	0.055**	0.057**	0.052*	111693

**Table 5:** Percentage of cases below or above the 90% three days ahead central IF grouped by adjusted  $R^2$ .

Adj. $R^2$	Below 90% IF			Above 90% IF			# cases
	irreg MF	irreg MF u-d	EWMA	irreg MF	irreg MF u-d	EWMA	
-20% - 0%	0.016**	0.016**	0.017**	0.045	0.045	0.045	5312
0 - 5%	0.017**	0.017**	0.018**	0.046**	0.046**	0.045**	23492
5 - 15%	0.021**	0.021**	0.023**	0.053**	0.053**	0.051	31150
15 - 30%	0.034**	0.037**	0.029**	0.059**	0.063**	0.056**	25799
30 - 45%	0.035**	0.040**	0.029**	0.057**	0.065**	0.052	12609
45 - 55%	0.049	0.062**	0.036**	0.069**	0.081**	0.062**	5041
55 - 65%	0.045	0.055	0.025**	0.052	0.062**	0.047	2900
65 - 75%	0.051	0.069**	0.025**	0.062*	0.079**	0.049	1602
75 - 85%	0.066*	0.090**	0.030**	0.052	0.069*	0.041	878
85 - 90%	0.087**	0.110**	0.054	0.084*	0.107**	0.066	335
90 - 95%	0.050	0.070	0.027	0.060	0.084*	0.030	299
95 - 100%	0.044	0.078	0.009*	0.044	0.078	0.035	115
all	0.028**	0.030**	0.024**	0.054**	0.057**	0.051	109532

The EWMA is a benchmark used in many financial institutions and showed to be conservative on the present data. Tables 4 and 5 show that the multi-factor model estimated with irregular returns brings an improvement to the plain EWMA in the central IF estimation, the exception being the upper bound three days ahead. Most of the classes have small but statistically significant deviations from the nominal tail percentage, from where we conclude that although the coverage is in general close to the nominal it is different. In general, the methods tended to be more conservative in the IF lower bound than in the upper bound, which means that an asymmetric predictive distribution could do better. However, using the semi-variance approach (results not shown here but available on the request) did not yield any better result.

As to the market index volatility proxy for the multi-factor, using the squared returns (irreg MF) performs better than using highs and lows (irreg MF u-d) for three days ahead, while the inverse occurs for one day ahead. As the **irreg MF u-d** seems to be too liberal for three days ahead, yielding an excessive number of cases above and below the IF when the adjusted  $R^2$  is above 45%, we recommend using the squared returns (weighted by EWMA) to estimate the factor volatility in the irregular returns multi-factor. The multi-factor model, estimated with irregular returns, has shown itself a useful tool to predict risk.

#### **4 – Conclusion**

In this paper we proposed the use of returns which are computed from prices irregularly spaced in time to estimate the multi-factor model for equity returns. The multi-factor model is a simple but efficient tool to explain cross-sectional covariance in equities returns. The model showed itself useful to estimate missing data as well as to provide interval forecasts for future returns. Furthermore, the use of irregular returns enables the estimation in cases where using only regular (daily) returns would not.

An empirical example with data from the 389 most liquid equities in the Brazilian Market confirmed the superiority of the multi-factor estimated with irregular returns over the traditional regular returns version, as well as two benchmark methods (the naïf and mimicking the return of the market index). Moreover, the market index showed itself significant to explain equities returns whereas the US\$/Brazilian Real exchange rate and the Brazilian inter-bank overnight interest rate did not.

#### **Acknowledgements**

This work was carried out in 2001 when both authors were working as consultants to Algorithmics Brazil and through Algorithmics Brazil to the Sao Paulo Exchange Market (BOVESPA) and its clearinghouse (CBLIC). The authors would like to thank seminar participants at the Graduate School of Economics/ Getulio Vargas Foundation, and conference attendants at the Computation in Economics and Finance 2002 and at the Forecasting Financial Markets 2003 for useful comments on previous versions of this work. The second author (Souza) greatly acknowledges the financial support received afterwards by FAPERJ, which enabled to write the academic version of the paper.

#### **References**

- Alexander, C. O. (1996), “Evaluating RiskMetrics as a risk measurement tool for your operation: what are its advantages and limitations?”, *Derivatives: Use, Trading and Regulation*, 2, 3, 277-284.
- Alexander, C. O. and Leigh, C.T. (1997), “On the covariance matrices used in Value-at-Risk models”, *Journal of Derivatives*, 4, 3, 50-62.
- Atchison, M.D., Butler, K.C. and Simonds, R.R. (1987), “Nonsynchronous security trading and market index autocorrelation”, *Journal of Finance* 42, 111-118.
- Bhandari, L. C. (1988), “Debt/Equity ratio and expected common stock returns: empirical evidence”, *Journal of Finance*, 43, 507-528.

- Brooks, R., Faff, R., Fry, T. and Gunn, L. (2001), "Censoring and its impact on beta risk estimation", *working paper*.
- Brooks, R., Faff, R.W., Fry, T. and Newton, E. (2002), "The selectivity corrected market model and heteroscedasticity in stock returns: yet another look at volume versus garch", *working paper*.
- Chan, L., Hamao, Y. and Lakonishok, J. (1991), "Fundamentals and stock returns in Japan", *Journal of Finance*, 46, 1739-1764.
- Chen, N-F, Roll, R. and Ross, S.A. (1986), "Economic forces and the stock market", *Journal of Business*, 59, 3, 383-403.
- Christoffersen, P.F. (1998), "Evaluating Interval Forecasts", *International Economic Review*, 39, 4, 841-862.
- Dempster, A.P., Laird, N. M. e D.B.Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B*, 39, pp.1-38.
- Dimson, E. (1979), "Risk measurement when shares are subject to infrequent trading", *Journal of Financial Economics* 7, 197-226.
- Dimson, E. and Marsh, P.R. (1983), "The stability of UK risk measures and the problem of thin trading", *Journal of Finance* 38, 753-783.
- Easley, D., Kiefer, N.M., O'Hara, M. and Paperman, J.B. (1996), "Liquidity, information, and infrequently traded stocks", *Journal of Finance* 51, 1405-1436.
- Engle, R.F. and Russell, J.R. (1998), "Autoregressive conditional duration: a new model for irregularly spaced transaction data", *Econometrica* 66, 1127-1162.
- Fama, E. and French, K. (1992), "The cross-section of expected stock returns", *Journal of Finance*, 47, 427-466.
- Fowler, D. and Rorke, C. (1983), "Risk measurement when shares are subject to infrequent trading: comment", *Journal of Financial Economics* 12, 279-283.
- Garman, M.B. and Klass, M.J. (1980), "On the estimation of security price volatilities from historical data", *Journal of Business*, 53, 1, 67-78.
- Gastineau, G. L. and Mark P. Kristzman, (1996), Dictionary of Financial Risk Management; Frank J. Fabozzi Associates.
- Jegadeesh, N. (1992), "Evidence of predictable behavior of security returns", *Journal of Finance*, 45, 881-898.
- Jorion, P. (1997), Value at Risk: a new benchmark for controlling market risk, Irwin Professional Publishing, Chicago.

- J P Morgan (1995), RiskMetrics – Technical Document, 3<sup>rd</sup> edition, Morgan Guaranty Trust Company, New York.
- Lo, A.W. and MacKinlay, A.C. (1988), “Stock market prices do not follow random walks: evidence from a simple specification test”, *Review of Financial Studies* 1, 41-66.
- Lo, A.W. and MacKinlay, A.C. (1990), “An econometric analysis of nonsynchronous trading”, *Journal of Econometrics* 45, 181-211.
- Marsh, P. (1979), “Equity rights issues and the efficiency of the UK stock market”, *Journal of Finance* 34, 839-862.
- Mei, J. (1993), “Explaining the cross-section returns via a multi-factor APT model”, *Journal of Financial and Quantitative Analysis*, 28, 3, 331-345.
- Neter, J., Kutner, M.H., Nachtsheim, C. and Wasserman, W. (1996), Applied Linear Regression Models, Irwin.
- Parkinson, M. (1980), “The extreme value method for estimating the variance of the rate of return”, *Journal of Business*, 53, 61-65.
- Reinganun, M. R. (1981), “Empirical tests of multi-factor pricing model. The Arbitrage Pricing Theory: some empirical results”, *Journal of Finance*, 36, 2, 313-321.
- Ross, S. (1976), “The arbitrage theory of capital asset pricing”, *Journal of Economic Theory*, 13, 341-360.
- Sharpe, W.F. (1964), “Capital Asset Prices: a theory of the market equilibrium under conditions of risk”, *Journal of Finance*, 19, 425-442.
- Scholes, M. and Williams, J. (1977), “Estimating betas from non-synchronous data”, *Journal of Financial Economics* 5, 309-327.
- Souza, L. and Veiga, A. (2001), “A comparison between the EM and the irregular returns multi-factor for missing data in emergent stock markets”, Algorithmics Brazil internal report.
- Subramanian, K.V. (2001), “Term structure estimation in illiquid markets”, *Journal of Fixed Income*, 11, 1, 77-86.

# ENSAIOS ECONÔMICOS DA EPGE

441. TESTING PRODUCTION FUNCTIONS USED IN EMPIRICAL GROWTH STUDIES – Pedro Cavalcanti Ferreira; João Victor Issler; Samuel de Abreu Pessoa – Março de 2002 - 28 págs.
442. THREE PRINCIPLES OF COMPETITIVE NONLINEAR PRICING – Frank H. Page Jr; Paulo K. Monteiro – Abril de 2002 – 44 págs.
443. O MECANISMO MONETÁRIO DE TRANSMISSÃO NA ECONOMIA BRASILEIRA PÓS-PLANO REAL – Marcelo Fernandes; Juan Toro – Abril de 2002 – 33 págs.
444. A ROBUST POVERTY PROFILE FOR BRAZIL USING MULTIPLE DATA SOURCES – Francisco H. G. Ferreira; Peter Lanjouw; Marcelo Neri – Abril de 2002 – 38 págs.
445. THE MISSING LINK: USING THE NBER RECESSION INDICATOR TO CONSTRUCT COINCIDENT AND LEADING INDICES OF ECONOMIC ACTIVITY – João Victor Issler; Farshid Vahid – Maio de 2002 – 31 págs.
446. SPECULATIVE ATTACKS ON DEBTS, DOLLARIZATION AND OPTIMUM CURRENCY AREAS – Aloisio Araujo; Marcia Leon – Maio de 2002 – 74 págs.
447. VINTAGE CAPITAL, DISTORTIONS AND DEVELOPMENT – Samuel de Abreu Pessoa; Rafael Rob – Maio de 2002 – 43 págs.
448. GOVERNMENT ACTIONS TO SUPPORT COFFEE PRODUCERS – AN INVESTIGATION OF POSSIBLE MEASURES FROM THE EUROPEAN UNION – Germán Calfat; Renato G. Flôres Jr. – Junho de 2002 – 30 págs.
449. A MODEL OF CAPITAL ACCUMULATION AND RENT-SEEKING – Paulo Barelli; Samuel de Abreu Pessoa – Junho de 2002 – 50 págs.
450. THE MISSING LINK: USING THE NBER RECESSION INDICATOR TO CONSTRUCT COINCIDENT AND LEADING INDICES OF ECONOMIC ACTIVITY – João Victor Issler; Farshid Vahid – Junho de 2002 – 29 págs.
451. LENDER LIABILITY IN THE CONSUMER CREDIT MARKET – Elisabetta Iossa; Giuliana Palumbo – Agosto de 2002 – 20 págs.
452. DECISION RULES AND INFORMATION PROVISION: MONITORING VERSUS MANIPULATION - Elisabetta Iossa; Giuliana Palumbo – Agosto de 2002 – 37 págs.
453. ON CERTAIN GEOMETRIC ASPECTS OF PORTFOLIO OPTIMISATION WITH HIGHER MOMENTS – Gustavo M. de Athayde; Renato G. Flôres Jr. – Setembro de 2002 – 21 págs.
454. MENSURANDO A PRODUÇÃO CIENTÍFICA INTERNACIONAL EM ECONOMIA DE PESQUISADORES E DEPARTAMENTOS BRASILEIROS - João Victor Issler; Tatiana Caldas de Lima Aché Pillar – Setembro de 2002 – 39 págs.
455. FOREIGN DIRECT INVESTMENT SPILLOVERS: ADDITIONAL LESSONS FROM A COUNTRY STUDY - Renato G. Flores Jr; Maria Paula Fontoura; Rogério Guerra Santos – Setembro de 2002 – 30 págs.

456. A CONTRACTIVE METHOD FOR COMPUTING THE STATIONARY SOLUTION OF THE EULER EQUATION - Wilfredo L. Maldonado; Humberto Moreira – Setembro de 2002 – 14 págs.
457. TRADE LIBERALIZATION AND THE EVOLUTION OF SKILL EARNINGS DIFFERENTIALS IN BRAZIL - Gustavo Gonzaga; Naércio Menezes Filho; Cristina Terra – Setembro de 2002 – 31 págs.
458. DESEMPENHO DE ESTIMADORES DE VOLATILIDADE NA BOLSA DE VALORES DE SÃO PAULO - Bernardo de Sá Mota; Marcelo Fernandes – Outubro de 2002 – 37 págs.
459. FOREIGN FUNDING TO AN EMERGING MARKET: THE MONETARY PREMIUM THEORY AND THE BRAZILIAN CASE, 1991-1998 - Carlos Hamilton V. Araújo; Renato G. Flores Jr. – Outubro de 2002 – 46 págs.
460. REFORMA PREVIDENCIÁRIA: EM BUSCA DE INCENTIVOS PARA ATRAIR O TRABALHADOR AUTÔNOMO - Samantha Taam Dart; Marcelo Côrtes Neri; Flavio Menezes – Novembro de 2002 – 28 págs.
461. DECENT WORK AND THE INFORMAL SECTOR IN BRAZIL – Marcelo Côrtes Neri – Novembro de 2002 – 115 págs.
462. POLÍTICA DE COTAS E INCLUSÃO TRABALHISTA DAS PESSOAS COM DEFICIÊNCIA - Marcelo Côrtes Neri; Alexandre Pinto de Carvalho; Hessia Guilherme Costilla – Novembro de 2002 – 67 págs.
463. SELETIVIDADE E MEDIDAS DE QUALIDADE DA EDUCAÇÃO BRASILEIRA 1995-2001 - Marcelo Côrtes Neri; Alexandre Pinto de Carvalho – Novembro de 2002 – 331 págs.
464. BRAZILIAN MACROECONOMICS WITH A HUMAN FACE: METROPOLITAN CRISIS, POVERTY AND SOCIAL TARGETS – Marcelo Côrtes Neri – Novembro de 2002 – 61 págs.
465. POBREZA, ATIVOS E SAÚDE NO BRASIL - Marcelo Côrtes Neri; Wagner L. Soares – Dezembro de 2002 – 29 págs.
466. INFLAÇÃO E FLEXIBILIDADE SALARIAL - Marcelo Côrtes Neri; Maurício Pinheiro – Dezembro de 2002 – 16 págs.
467. DISTRIBUTIVE EFFECTS OF BRAZILIAN STRUCTURAL REFORMS - Marcelo Côrtes Neri; José Márcio Camargo – Dezembro de 2002 – 38 págs.
468. O TEMPO DAS CRIANÇAS - Marcelo Côrtes Neri; Daniela Costa – Dezembro de 2002 – 14 págs.
469. EMPLOYMENT AND PRODUCTIVITY IN BRAZIL IN THE NINETIES - José Márcio Camargo; Marcelo Côrtes Neri; Maurício Cortez Reis – Dezembro de 2002 – 32 págs.
470. THE ALIASING EFFECT, THE FEJER KERNEL AND TEMPORALLY AGGREGATED LONG MEMORY PROCESSES - Leonardo R. Souza – Janeiro de 2003 – 32 págs.
471. CUSTO DE CICLO ECONÔMICO NO BRASIL EM UM MODELO COM RESTRIÇÃO A CRÉDITO - Bárbara Vasconcelos Boavista da Cunha; Pedro Cavalcanti Ferreira – Janeiro de 2003 – 21 págs.
472. THE COSTS OF EDUCATION, LONGEVITY AND THE POVERTY OF NATIONS - Pedro Cavalcanti Ferreira; Samuel de Abreu Pessoa – Janeiro de 2003 – 31 págs.

473. A GENERALIZATION OF JUDD'S METHOD OF OUT-STEADY-STATE COMPARISONS IN PERFECT FORESIGHT MODELS - Paulo Barelli; Samuel de Abreu Pessoa – Fevereiro de 2003 – 7 págs.
474. AS LEIS DA FALÊNCIA: UMA ABORDAGEM ECONÔMICA - Aloísio Pessoa de Araújo – Fevereiro de 2003 – 25 págs.
475. THE LONG-RUN ECONOMIC IMPACT OF AIDS - Pedro Cavalcanti G. Ferreira; Samuel de Abreu Pessoa – Fevereiro de 2003 – 30 págs.
476. A MONETARY MECHANISM FOR SHARING CAPITAL: DIAMOND AND DYBVG MEET KIYOTAKI AND WRIGHT – Ricardo de O. Cavalcanti – Fevereiro de 2003 – 16 págs.
477. INADA CONDITIONS IMPLY THAT PRODUCTION FUNCTION MUST BE ASYMPTOTICALLY COBB-DOUGLAS - Paulo Barelli; Samuel de Abreu Pessoa – Março de 2003 – 4 págs.
478. TEMPORAL AGGREGATION AND BANDWIDTH SELECTION IN ESTIMATING LONG MEMORY - Leonardo R. Souza - Março de 2003 – 19 págs.
479. A NOTE ON COLE AND STOCKMAN - Paulo Barelli; Samuel de Abreu Pessoa – Abril de 2003 – 8 págs.
480. A HIPÓTESE DAS EXPECTATIVAS NA ESTRUTURA A TERMO DE JUROS NO BRASIL: UMA APLICAÇÃO DE MODELOS DE VALOR PRESENTE - Alexandre Maia Correia Lima; João Victor Issler – Maio de 2003 – 30 págs.
481. ON THE WELFARE COSTS OF BUSINESS CYCLES IN THE 20TH CENTURY - João Victor Issler; Afonso Arinos de Mello Franco; Osmani Teixeira de Carvalho Guillén – Maio de 2003 – 29 págs.
482. RETORNOS ANORMAIS E ESTRATÉGIAS CONTRÁRIAS - Marco Antonio Bonomo; Ivana Dall'Agnol – Junho de 2003 – 27 págs.
483. EVOLUÇÃO DA PRODUTIVIDADE TOTAL DOS FATORES NA ECONOMIA BRASILEIRA: UMA ANÁLISE COMPARATIVA - Victor Gomes; Samuel de Abreu Pessoa; Fernando A . Veloso – Junho de 2003 – 45 págs.
484. MIGRAÇÃO, SELEÇÃO E DIFERENÇAS REGIONAIS DE RENDA NO BRASIL - Enestor da Rosa dos Santos Junior; Naércio Menezes Filho; Pedro Cavalcanti Ferreira – Junho de 2003 – 23 págs.
485. THE RISK PREMIUM ON BRAZILIAN GOVERNMENT DEBT, 1996-2002 - André Soares Loureiro; Fernando de Holanda Barbosa - Junho de 2003 – 16 págs.
486. FORECASTING ELECTRICITY DEMAND USING GENERALIZED LONG MEMORY - Lacir Jorge Soares; Leonardo Rocha Souza – Junho de 2003 – 22 págs.
487. USING IRREGULARLY SPACED RETURNS TO ESTIMATE MULTI-FACTOR MODELS: APPLICATION TO BRAZILIAN EQUITY DATA - Álvaro Veiga; Leonardo Rocha Souza – Junho de 2003 – 26 págs.